SOME RANKING AND SELECTION CRITERIA FOR DETERMINING SAMPLE SIZES IN TWO DIFFERENT MODELS FOR A DRUG COMBINATION PROBLEM*

by

F. N. Lin and M. Sobel

Technical Report No. 197

January 1973

University of Minnesota

Minneapolis, Minnesota

*This work was partially supported by Grant DA-ARO-D-31-124-72-6187 at the University of Minnesota and partially by contract NIH-E-71-2180 at the University of California in San Francisco.

1. Introduction.

We consider some simple models for combining pairs of chemicals in each test animal and apply some ideas and methods from the area of ranking and selection problems to determine the optimal sizes of the various test groups that are necessary to satisfy prescribed probability requirements. Our results are asymptotic since we use normal distribution theory for a binomial problem with large sample sizes.

Four chemicals (principally food additives) are to be studied for tumorigenic effects on a single species of test animal. Our goal is to classify each of the four chemicals as tumorigenic or not and at the same time classify the pairs of chemicals as being tumorigenic or not. In this initial study we use a simple model that allows us to test only pairs of chemicals (without testing each chemical separately) and still satisfy three lower bound requirements on the probability of a correct classification. The one-at-a-time method is briefly looked at for the purpose of comparisons but this clearly gives no information about the possible results for pairs of chemicals and hence this comparison is not very fair or meaningful, except to estimate the cost in total sample size of getting additional information about pairs of chemicals.

These three probability requirements all refer to different configurations in the parameter space (i.e., states of nature) and hence we are not concerned by the fact that these three different compound events (all dealing with a correct classification) are not independent.

We make use of a control group where none of these four chemicals are used and all of our comparisons are with respect to the results in this common control group. If we repeat the experiment on other sets of four chemicals then after several repetitions the control group can either be

eliminated or, even if not eliminated, can be treated like a group with a known p-value, namely $\underline{p} = 1 - \underline{q}$. Hence we consider in Section 6 the corresponding problem in which no observations are taken on the control; the answers are sought for different values of $\underline{p}$ and comparisons are made.

The particular plan or procedure treated in this paper considers all $\binom{4}{2} = 6$ combinations of the four chemicals simultaneously and without the benefit of any sequential plan. In Section 8 we briefly consider some plans for future research that cut down on the number of combinations and involve 2-stage procedures that are sequential in nature. These will result in a substantial saving in the total number of observations needed.

2. Formulation of the Problem With a Control.

Four chemicals C_i $(i = 1,2,3,4)$ are given and we determine the number N of test animals for each of the $\binom{4}{2} = 6$ pairs of chemicals. In addition we put kN test animals in a control group where none of the four chemicals are used. For approximate calculations we shall be interested in $k = 3$ (or $k = 3.5$); for exact calculations we later determine k as well as N. We assume that the final result for N and kN will be moderately large so that the normal approximation theory can be used with negligible error.

Let p_i denote the probability that chemical C_i $(i = 1,2,3,4)$ produces one or more tumors in a test animal drawn at random as a representative of the population. Let $p_{i,j}$ denote the corresponding probability of producing tumors when both C_i and C_j $(i \neq j;\ i,j = 1,2,3,4)$ are used in the same test animal. Two models of interest are:

$$(1.1) \qquad p_{i,j} = 1 - (1-p_i)(1-p_j) + \delta_{i,j}^{(1)} - pq \qquad \underline{\text{Model 1}}.$$

$$(1.2) \qquad p_{i,j} = \text{Max}(p_i, p_j) + \delta_{i,j}^{(2)} \qquad \underline{\text{Model 2}}.$$

Here $\delta_{ij}^{(1)} - \underline{pq}$ and $\delta_{ij}^{(2)}$ denote 'interaction' between C_i and C_j under Models 1 and 2, respectively; we use δ_{ij} below without superscript to apply to both models. Both models have the property that when $\delta_{ij} = 0$ and $p_i = p_j = \underline{p}$, the value of p_{ij} is also $\underline{p}$. Hence $\underline{p}$ can be regarded as the background rate for animals to develop tumors when the chemicals administered have no effect singly or in combination; thus we are assuming in (1.1) that $p_i \geq \underline{p}$ for each i. Under Model 1 when the interaction $\delta_{ij}^{(1)} - \underline{pq} = 0$ we have the usual type of statistical independence. Under Model 2 with $\delta_{ij}^{(2)} = 0$ one of the chemicals C_i and C_j dominated the other chemical it was paired with. In our illustrations below we compare $\delta_{ij}^{(1)} = \delta_0$ with $\delta_{ij}^{(2)} = \delta_0$, since they are comparable at $\delta_0 = 0$. We do not claim to have experimental data to justify either of these models but many of our derivations (with at most slight changes) can be used for both models. Both models have been used in the area of drug combinations by other authors (cf. Bliss [1], Finney [2] and Plackett and Hewlitt [3]).

Let X_{ij} denote the number of tumorigenic test animals (i.e., the number of test animals with at least one tumor) arising from the set(of size N) given the pair of chemicals C_i and C_j $(i \neq j;\ i,j = 1,2,3,4)$. Define

$$
\begin{aligned}
X_1 &= X_{12} + X_{13} + X_{14} \\
X_2 &= X_{12} + X_{23} + X_{24} \\
X_3 &= X_{13} + X_{23} + X_{34} \\
X_4 &= X_{14} + X_{24} + X_{34} \ .
\end{aligned} \tag{1.3}
$$

It should be noted that X_i represents the effect of chemical C_i $(i = 1,2,3,4)$ only in very special states of nature (which we spell out below); in particular, the X_i are not used to obtain general confidence intervals for p_i under

either of our two models since they overestimate the value of p_i in both cases. Let X_0 denote the number of tumorigenic test animals in the control group of size kN. Our procedure is to assert that chemical C_i is tumorigenic if (and only if)

$$\frac{X_i}{3} - \frac{X_0}{k} > Nc, \tag{1.4}$$

where c is a positive fraction yet to be determined.

A chemical is called tumorigenic if its p-value is at least $\overline{p}$ and it is called non-tumorigenic if its p-value is at most $\underline{p}$, where $\underline{p} \leq \overline{p}$; we take $\underline{p} = .15$ and $\overline{p} = .30$ in our numerical illustrations below. If the p-value is between $\underline{p}$ and $\overline{p}$ then either assertion (or no assertion) can be regarded as acceptable (i.e., as correct), but if $p \leq \underline{p}$ or $p \geq \overline{p}$ then only one assertion is correct. Note that we are using an indifference zone approach, namely $(\underline{p}, \overline{p})$, to calling a single chemical tumorigenic and that this indifference zone is the same for all chemicals and for both models. In the discussion below we also use the symbol Δ for the length of the indifference zone, i.e., $\Delta = \overline{p} - \underline{p}$.

Our procedure for the combination aspect is to assert that the pair (C_i, C_j) is tumorigenic if (and only if)

$$X_{i,j} - \frac{X_0}{k} > Nc' \tag{1.5}$$

where c' is another positive factor to be determined and k, N are as defined above.

The symbol CD stands for a correct decision and indicates the event that a correct classification (tumorigenic or non-tumorigenic) was made simultaneously on the individual chemicals and also on the pairs. Let CD_1 denote the joint event that the four individual chemicals are classified correctly and let CD_2 denote the corresponding joint event for the six pairs of chemicals.

We pay special attention to three states of nature, ω_0, ω_1 and ω_2, given under both models by

$$
\begin{aligned}
&\omega_0\colon\ p_0 = p_1 = p_2 = p_3 = p_4 = \underline{p};\ \delta_{ij} = 0 \quad \text{for all pairs } (i, j)\\
&\omega_1\colon\ p_0 = p_1 = p_2 = p_3 = \underline{p};\ p_4 = \overline{p};\ \delta_{ij} = 0 \quad \text{for all pairs } (i, j)\\
&\omega_2\colon\ p_0 = p_1 = p_2 = p_3 = p_4 = \underline{p};\ \delta_{34} = \Delta;\ \delta_{ij} = 0 \text{ for all pairs } (i,j) \neq (3,4).
\end{aligned}
\tag{1.6}
$$

In ω_0 the values of p_{ij} are the same under both models. In ω_1 the value of p_{i4} $(i = 1,2,3)$ is $p_a = \overline{p} - \underline{p}\Delta = \underline{p} + \underline{q}\Delta$ (between $\underline{p}$ and $\overline{p}$) under Model 1, and it is $\overline{p}$ under Model 2; for other pairs (i, j) the values are the same under both models, namely $\underline{p}$. In ω_2 the value of p_{34} is $\underline{p} + \Delta$ $(\Delta \geq 0)$ and the value of p_{ij} is $\underline{p}$ for all pairs $(i,j) \neq (3,4)$ and these hold for both models.

At this point we mention an assumption to be used later; let $\overline{q} = 1 - \overline{p}$, $\underline{q} = 1 - \underline{p}$ (used above) and $q_a = 1 - p_a$. After (3.9), (4.9) and (5.3) we use the assumption that

$$0 < \underline{pq} < p_a q_a < \overline{p}\,\overline{q} < 3\underline{pq}. \tag{1.7}$$

This assumption is easily seen to hold in our numerical illustration where $\underline{p} = .15$, $\overline{p} = .30$ and hence $p_a = \underline{p} + \underline{q}\Delta = .2775$.

The basic probability requirements that we wish to satisfy can now be stated in terms of (1.6) and three preassigned constants P_0^*, P_1^*, P_2^*, which are all less than one. We want to find a quadruple (N, c, c', k) with the smallest value of $(6+k)N$ and such that we simultaneously satisfy all of the following inequalities,

$$P_0 = P\{CD_1|\omega_0\} \geq P_0^*, \tag{1.8}$$

$$P_1 = P\{CD_1|\omega_1\} \geq P_1^*, \tag{1.9}$$

$$P_2 = P\{CD_2|\omega_2\} \geq P_2^*. \tag{1.10}$$

In our illustration below we take $P_0^* = .99$, $P_1^* = .90$ and $P_2^* = .75$ and .90

Remark 1.

If our main goal is to form conclusions about the individual chemicals then we want to impose the P_1^* condition in (1.9) and treat P_2 on the left side of (1.10) as an important additional property of our procedure that should be calculated.

Remark 2.

For purposes of comparing our procedure with the one-at-a-time procedure it is necessary to disregard the requirement (1.10) but even in this case we would find the value of c' that maximizes P_2 and the resulting value of P_2. If the results of the combinations form the main part of an important part of the goal then the one-at-a-time procedure cannot be used and this comparison is meaningless. As mentioned earlier, a comparison with the one-at-a-time procedure is important from the point of view of the overall cost of getting information about the individual chemicals only.

Remark 3.

In each of the above requirements it will be seen later in our example that some decisions can be added with a negligible change in the joint probability level. Thus for P_0 we might like to change CD_1 to CD. For P_1 we might like to add the decisions that the pairs (1,2), (1,3) and (2,3) are non-tumorigenic without altering the joint probability level, P_1^*. For P_2 we might like to add the decisions that the four chemicals C_1, C_2, C_3, and C_4 are separately non-tumorigenic without altering the joint probability level, P_2^*, i.e., change CD_2 to CD. These omitted decisions will be shown later to have a negligible effect in our example but their effect has not been evaluated or bounded in the general case. Since we omitted them in our

requirements above, they do not affect the generality of our solution. In a more positive vein it should be noted that as it stands now with CD_1 the requirement in (1.9) is more comparable with the corresponding power calculation for the one-at-a-time procedure. Ease of calculation is also an issue that favors the formulations given in (1.8), (1.9), and (1.10).

Remark 4.

We allow our solution in N to be a real number and use the same symbol for N whether it is an integer or not. If it is not an integer then the next larger integer will suffice and, if desired, a randomized solution between two consecutive integers can also be used. Allowing N to be decimal, we do assume that equality will be obtained in at least one of these three requirements, but we do not expect to be able to attain, for example, $P_0 = P_1$ when $P_0^* = P_1^*$. In our illustration we set $P_1 = P_1^* = .9$ and find that P_0 is extremely close to one. For checking the inequality in (1.8) it turns out that we can use existing tables of the integral

$$(1.11) \qquad I_{k'}(h, \rho') = \int_{-\infty}^{\infty} \Phi^{k'-1}\left(\frac{x\sqrt{\rho'} + h}{\sqrt{1-\rho'}} \right) \varphi(x)dx$$

where $\Phi(x)$ and $\varphi(x)$ are the standard normal c.d.f. and density, respectively (cf., [ ] or [ ]; note that the latter use k' instead of $k'-1$ on the right side of (1.11)). The requirement (1.9) (resp., (1.10) requires the use of quadrature (say, Gauss-Hermite) to evaluate the double (resp., single) integral that it gives rise to.

3. Derivation of Integral Expressions for p_i $(i = 0,1,2)$.

Under both models in configuration ω_0 the means of the X_i are $\mu_i = 3N\underline{p}$ $(i = 1,2,3,4)$ and the mean of the control X_0 is $\mu_0 = kN\underline{p}$. The variances are ${\sigma_1}^2 = 3N\underline{pq}$ $(i = 1,2,3,4)$ and for the control ${\sigma_0}^2 = kN\underline{pq}$. Let ρ denote the common correlation between X_i and X_j $(i \neq j;\ i,j = 1,2,3,4)$;

under ω_0 the value of ρ is easily seen to be $1/3 > 0$. Using asymptotic normality for large N we first write P_0 as

$$(3.1) \qquad P_0 = P\left\{ \frac{X_i}{3} - \frac{X_0}{k} < cN \ (i = 1,2,3,4) \middle| \omega_0 \right\}$$

$$\tilde{=} P\{Y_i < Y_0\sqrt{3/k} + c\sqrt{3N/pq} \ (i = 1,2,3,4) | \omega_0\}$$

where the Y_i $(i = 0,1,2,3,4)$ are standardized normal chance variables but are not independent. To evaluate (3.1) we introduce five independent standardized normal chance variables (all mutually independent of Y_0) defined by

$$(3.2) \qquad Y_i = Z_i\sqrt{1-\rho} - Z_0\sqrt{\rho} \quad (i = 1,2,3,4)$$

and another (sixth) standard normal chance variable defined by

$$(3.3) \qquad V = \frac{Z_0\sqrt{\rho} + Y_0\sqrt{3/k}}{\sqrt{\rho+(3/k)}} \ .$$

Since the correlation structure of the Y_i in (3.2) is the same as for the Y_i (and as for the X_i) $(i = 1,2,3,4)$ in (3.1), we can substitute (3.2) and (3.3) in (3.1). This gives for $\rho = 1/3$ under both models

$$(3.4) \qquad P_0 = P\{Z_i < v\sqrt{\frac{\rho+3/k}{1-\rho}} + c\sqrt{\frac{3N}{pq(1-\rho)}} \quad (i = 1,2,3,4)\}$$

$$= \int_{-\infty}^{\infty} \Phi^4(v\sqrt{\frac{k+9}{2k}} + 3c\sqrt{\frac{N}{2pq}} \ )\varphi(v)dv.$$

Comparing with the standard form (1.13) we find that $k' = 5$,

$$(3.5) \qquad \rho' = \frac{\rho + 3/k}{1 + 3/k} = \frac{k+9}{3(k+3)} \ , \ \frac{h}{c} = \sqrt{\frac{3Nk}{pq(k+3)}} \ ,$$

where $h > 0$ does not depend on ρ, and $\rho' < 1$ does not depend on p or P.

To obtain an integral expression for P_1 we first note that under Model I in configuration ω_1 the means, variances and correlations are

$$\mu_0 = kN\underline{p};\ \mu_1 = \mu_2 = \mu_3 = N(2\underline{p}+p_a);\ \mu_4 = 3Np_a,$$

$$\sigma_0^2 = kN\underline{pq};\ \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = N(2\underline{pq} + p_aq_a);\ \sigma_4^2 = 3Np_aq_a,$$

$$\rho_{12} = \rho_{13} = \rho_{23} = \frac{\underline{pq}}{2\underline{pq} + \overline{pq}} > 0, \tag{3.6}$$

$$\rho_{14} = \rho_{24} = \rho_{24} = \sqrt{\frac{p_aq_a}{3(2\underline{pq} + p_aq_a)}} > 0,$$

where ρ_{ij} is the correlation between X_i and X_j. The only change needed for Model 2 is that we replace p_aq_a everywhere it appears by $\overline{pq}$; this remark remains valid through the derivation of P_1 below in (3.10). Defining Y_i $(i = 0,1,2,3,4)$ as in (3.1) and using asymptotic normality for large N we can write P_1 as

$$P_1 = P\{\frac{X_4}{3} - \frac{X_0}{k} \geq cN,\ \frac{X_i}{3} - \frac{X_0}{k} < cN(i = 1,2,3)\,|\,\omega_1\} \tag{3.7}$$

$$\cong P\{Y_4 \geq Y_0\sqrt{\frac{3\underline{pq}}{kp_aq_a}} + \frac{(c-\Delta')\sqrt{3N}}{\sqrt{\overline{pq}}},\ Y_i \leq \frac{Y_0\sqrt{2\underline{pq}/k} + (3c-\Delta')\sqrt{N}}{\sqrt{2\underline{pq} + p_aq_a}}(i=1,2,3)\}$$

where $\Delta' = p_a - \underline{p} = \underline{q}\Delta > 0$. We introduce five independent standard normal chance variables Z_i $(i = 0,1,2,3,4)$ defined by

$$Y_i = Z_i\sqrt{1 - \rho_{12}} - Z_0\sqrt{\rho_{12}}\ (i = 1,2,3)$$
$$Y_4 = Z_4\sqrt{1 - \rho_1} - Z_0\sqrt{\rho_1} \tag{3.8}$$

where ρ_1 is defined by the relation $\sqrt{\rho_{12}\rho_1} = \rho_{14}$ or

$$\rho_1 = \frac{\rho_{14}^2}{\rho_{12}} = \frac{p_aq_a}{3\underline{pq}}. \tag{3.9}$$

Hence to keep Y_4 real we have to assume that $\rho_1 \leq 1$ or $\rho_{14}^2 \leq \rho_{12}$; this follows from (1.7). From (3.8) and (3.7) we have, using (3.6) and (3.9), for Model 1

$$(3.10)\qquad P_1 \cong P\{Z_4 > \frac{Z_0\sqrt{\rho_1 p_a q_a} + Y_0\sqrt{3\underline{pq}/k} + (c-\Delta')\sqrt{3N}}{\sqrt{(1-\rho_1)p_a q_a}},$$

$$Z_i < \frac{Z_0\sqrt{\rho_{12}(2\underline{pq} + p_a q_a)} + Y_0\sqrt{9\underline{pq}/k} + (3c-\Delta')\sqrt{N}}{\sqrt{(1-\rho_{12})(2\underline{pq} + p_a q_a)}} \quad (i = 1,2,3)\}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [1 - \Phi(\frac{zp_a q_a + 3y\underline{pq}/\sqrt{k} - 3(\Delta'-c)\sqrt{N\underline{pq}}}{\sqrt{p_a q_a(3\underline{pq} - p_a q_a)}})]$$

$$\cdot \Phi^3(\frac{z\sqrt{\underline{pq}} + 3y\sqrt{\underline{pq}/k} + (3c-\Delta')\sqrt{N}}{\sqrt{\underline{pq} + p_a q_a}})\varphi(z)\varphi(y)dzdy.$$

Thus we note that P_1 is given by a double integral and require numerical quadrature. If we restrict c to the open interval $(\Delta'/3, \Delta')$ then $P_1 \to 1$ as $N \to \infty$ and if either $c < \Delta'/3$ or $c > \Delta'$ then $P_1 \to 0$ as $N \to \infty$. Similar results for any c hold for P_0 in (3.4) and for certain values of c' hold for P_2 in (3.12). Hence for any c in this open interval and any $P_i^* < 1$ $(i = 1,2,3)$ there must exist a smallest integer N or a pair (k, N) with the smallest value of $(6+k)N$ that satisfies (1.8) and (1.9) or (1.8), (1.9) and (1.10).

For P_2 we use the seven independent quantities X_{ij} and X_0 and asymptotic normality with large N to obtain the corresponding integral form. It is easily seen that in ω_2 the value of $p_{3,4}$ is $\bar{p}$ under both models. Let $\Delta = \bar{p} - \underline{p}$ as before. Then under either model we obtain

$$(3.11)\qquad P_2 = P\{X_{3,4} - \frac{X_0}{k} \geq c'N,\ X_{i,j} - \frac{X_0}{k} < c'N \text{ (for all } (i,j) \neq (3,4)|\omega_2\}$$

$$\cong P\{Y_{3,4} > \frac{Y_0\sqrt{\underline{pq}/k} + (c'-\Delta)\sqrt{N}}{\sqrt{\underline{pq}}},\ Y_{i,j} < \frac{Y_0\sqrt{\underline{pq}/k} + c'\sqrt{N}}{\sqrt{\underline{pq}}}, (i,j) \neq (3,4)\}$$

where the seven Y's are standardized normal chance variables as before and are mutually independent. Then the resulting expression is a single integral given by

$$(3.12)\qquad P_2 = \int_{-\infty}^{\infty} [1 - \Phi(\frac{y\sqrt{pq/k} - (\Delta - c')\sqrt{N}}{\sqrt{pq}})]\Phi^5(\frac{y\sqrt{pq/k} + c'\sqrt{N}}{\sqrt{pq}})\varphi(y)dy.$$

This integral and the double integral in (3.10) both need numerical quadrature for exact evaluation; in the next section we show how to obtain approximate evaluations for each of them without any quadrature. The error of these approximations is shown to be small for large N. If $0 < c' < \Delta$ then both factors in the integrand in (3.12) approach one and hence there must exist a smallest N (or a pair (k, N) with the smallest value of $(6+k)N$) that satisfies (1.10).

We wish to show that for many reasonable values of $\underline{p}$ and $\overline{P}$ the requirement (1.9) will determine N, k and c and (1.10) will determine c'; in these cases the requirement (1.8) is not used at all except that we use (1.11) to check that (1.8) is satisfied after finding N, k, c and c'.

4. An Approximate Solution for Equality in (1.9).

For $1 - P_1^*$ close to zero we replace $\Phi^3(x)$ in (3.10) for x large by

$$(4.1)\qquad \Phi^3(x) = [1 - \Phi(-x)]^3 \sim 1 - 3\Phi(-x),$$

where '$\sim$' denotes an approximation for large N (which will imply that x is large). It is easily seen that a strict inequality holds in (4.1) for all x. Furthermore, for x and v both large we use (4.1) and obtain

$$(4.2)\qquad 1 - \Phi^3(x)[1 - \Phi(-v)] \sim \Phi(-v) + 3\Phi(-x),$$

if both x and $v \to \infty$ as $N \to \infty$. We assume (and check later) that the desired solution for our illustrative example involves an equality in (1.9) and a strict inequality in (1.8). Taking compliments in the last expression of (3.10) and using (4.2), we obtain

$$(4.3)\qquad 1 - P_1^* \sim \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\Phi\left(\frac{z\sqrt{p_a q_a} + 3y\underline{pq}/\sqrt{k p_a q_a} - 3(\Delta'-c)\sqrt{N\underline{pq}/p_a q_a}}{\sqrt{3\underline{pq} - p_a q_a}} \right.$$

$$\left. + 3\Phi\left(\frac{-z\sqrt{\underline{pq}} - 3y\sqrt{\underline{pq}/k} - (3c-\Delta')\sqrt{N}}{\sqrt{\underline{pq} + p_a q_a}} \right)]\varphi(z)\varphi(y)dzdy$$

$$= \Phi\left(\frac{-(\Delta'-c)\sqrt{3N}}{\sqrt{p_a q_a + 3\underline{pq}/k}} \right) + 3\Phi\left(\frac{-(3c-\Delta')\sqrt{N}}{\sqrt{p_a q_a + (2+\frac{9}{k})\underline{pq}}} \right).$$

Taking complements once again above gives the final form

$$(4.4)\qquad \Phi\left(\frac{(\Delta'-c)\sqrt{3N}}{\theta_1} \right) - 3\Phi\left(\frac{-(3c-\Delta')\sqrt{N}}{\theta_2} \right) \sim P_1^*,$$

where $\theta_1 = \theta_1(k)$ and $\theta_2 = \theta_2(k)$ are the denominator radicals in the last line of (4.3), in the order presented.

Here for given $\underline{p}$ and p_a (and $\Delta' = p_a - \underline{p}$) we want to find the pair N, k (and the corresponding c value) such that $(6+k)N$ is minimized and the inequality holds in (4.4). This result can then be used as a first approximation to the solution of the original condition, which involves the double integral in (3.10).

A first approximation to (4.4) is obtained by solving the pair of equations for some fixed k(say $k = 3$)

$$(4.5)\qquad \Phi\left(\frac{(\Delta'-c)\sqrt{3N}}{\theta_1} \right) = \frac{1 + P_1^*}{2},$$

$$(4.6)\qquad 3\Phi\left(\frac{-(3c-\Delta')\sqrt{N}}{\theta_2} \right) = \frac{1 - P_1^*}{2}.$$

If we let $\lambda_1 > 0$ and $\lambda_2 < 0$ denote the standard normal percentiles corresponding to $(1+P_1^*)/2$ and $(1-P_1^*)/6$, respectively, then the solution of (4.5) and (4.6) is easily seen to be

$$(4.7)\qquad N = \left(\frac{\theta_1\lambda_1\sqrt{3} - \theta_2\lambda_2}{2\Delta'} \right)^2 ; \quad c = \frac{\Delta'(\theta_2\lambda_2 - \theta_1\lambda_1/3)}{\theta_2\lambda_2 - \theta_1\lambda_1\sqrt{3}}.$$

We note that this value of c is between $\Delta'/3$ and Δ' but we will not use this formula for c in our algorithm. Although N changes in later corrections, the value of $(6+k)N$ changes by little in subsequent corrections.

Using the value of $N = N(k)$ from (4.7) as fixed, we wish to find the value of c that maximizes the left side of (4.4) and the value of k that minimizes $(6+k)N$. Then with these values of c and k in (4.4) we try to decrease N by integer amounts and still have '$\geq$' in (4.4). If we can decrease N then we revise the values of c and k in the same manner as before. This algorithm appears to converge fairly rapidly. A final check in (4.4) with a net of (c, k) pairs and a few consecutive N values then gives us a solution N_0 for (4.4). This is easily seen to be an upper bound on the correct answer since the inequality '$>$' holds in (4.1). Numerical quadrature on the double integral in (3.10) is then needed for a net of (c, k) pairs and a few consecutive values of $N \leq N_0$ to get the final solution for (3.10) and (1.9).

To get the value of c that maximizes the left side of (4.4) we treat N and k as fixed in (4.4) and differentiate the left side of (4.4) with respect to c. Setting the result equal to zero, we obtain

$$(4.8) \qquad \frac{1}{\theta_1}\varphi\left(\frac{(\Delta'-c)\sqrt{3N}}{\theta_1}\right) = \frac{3\sqrt{3}}{\theta_2}\varphi\left(\frac{-(3c-\Delta')\sqrt{N}}{\theta_2}\right);$$

this reduces to a quadratic equation in c, namely

$$(4.9) \qquad \frac{(3c-\Delta')^2}{\theta_2^2} - \frac{3(\Delta'-c)^2}{\theta_1^2} - \frac{2}{N}\ln\left(\frac{3\theta_1\sqrt{3}}{\theta_2}\right) = 0,$$

where $\ln$ denotes the natural logarithm.

It is easy to check that if p_a is closer to $1/2$ than $\underline{p}$ (i.e., if $p_a q_a > \underline{pq}$) then (regardless of the value of k) the coefficient of c^2 in (4.9) is positive and the value of the quadratic is negative at $c = \Delta'/3$.

It follows that there is one root $< \Delta'/3$ and hence we can restrict our attention to the larger root, which is given by

$$(4.10) \qquad c = \frac{-\Delta'(\theta_2^2 - \theta_1^2) + \theta_1\theta_2\sqrt{\frac{4}{3}(\Delta')^2 + \frac{2}{3N}(3\theta_1^2 - \theta_2^2)\ln(3\theta_1\sqrt{3} / \theta_2)}}{3\theta_1^2 - \theta_2^2} .$$

For large N this value of c will be in the interval $(\Delta'/3, \Delta')$ and, because the second derivative of (4.4) is easily shown to be negative, it must be a maximizing value of c.

To find the approximate value of k that minimizes $(6+k)N$ for $N = N(k)$ given by (4.7) we differentiate $(6+k)N$ with respect to k and obtain

$$(4.11) \qquad k = \left[\frac{-(1 + \frac{6}{k})\frac{\partial N}{\partial k}}{N}\right] = \frac{-N}{(1 + \frac{6}{k})\frac{\partial N}{\partial k}} .$$

We can get $\partial N/\partial k$ for the right side of (4.11) either by differentiating the first expression in (4.7), which does not depend on c, or by differentiating the left side of (4.4) and treating c as a constant. [A search for the exact k-value can then more easily be made by trial and error.] Both of these are approximate and lead to algorithms that converge fairly rapidly; the two results for approximating k are, respectively,

$$(4.12) \qquad k = \frac{3\sqrt{3}\,(1 + \frac{6}{k})pq(\theta_1\lambda_1\sqrt{3} - \theta_2\lambda_2)(\theta_2\lambda_1 - \theta_1\lambda_2\sqrt{3})}{4(\Delta')^2\theta_1\theta_2\, N} ,$$

$$(4.13) \qquad k = \frac{3pq(1 + \frac{6}{k})}{\theta_1^2}\left[\frac{(\frac{\Delta'-c}{\theta_1})\varphi(\frac{(\Delta'-c)}{\theta_1}\sqrt{3N}) + 3(\frac{\theta_1}{\theta_2})^2\sqrt{3}(\frac{3c-\Delta'}{\theta_2})\varphi(\frac{(3c-\Delta')\sqrt{N}}{\theta_2}}{(\frac{\Delta'-c}{\theta_1})\varphi(\frac{(\Delta'-c)}{\theta_1}\sqrt{3N}) + \sqrt{3}(\frac{3c-\Delta'}{\theta_2})\varphi(\frac{(3c-\Delta')\sqrt{N}}{\theta_2})}\right.$$

A good initial guess k_0 for k for both (4.12) and (4.13) is to set the entire square bracket in (4.13) equal to one, since $\varphi((3c-\Delta')\sqrt{N}/\theta_2)$ is

much smaller than $\varphi((\Delta'-c)\sqrt{3N} / \theta_1)$. This gives a quadratic equation in k and has the simple solution

$$(4.14) \qquad k_0 = 3\sqrt{\frac{2pq}{p_a q_a}} \quad .$$

For Model 1, if $\underline{p} = .15$, $\overline{p} = .30$ (hence $p_a = .2775$) and $P_1^* = .90$, then $\lambda_1 = 1.6450$, $\lambda_2 = -2.1282$, $\Delta' = p_a - \underline{p} = .2775 - .15 = .1275$ and the initial guess for k_0 by (4.14) is $k_0 = 3.4$. By (4.7) with $k_0 = 3.4$ the initial guess for N is 188 and $c = .08$; by (4.10) the estimate of c is .09. Then the value of c changes very little and after some searching by numerical quadrature on (3.10) we obtain $N = 188$, $c = .089$, and $k = 2.97$; this gives $P_1 = .9005$ and results in a total number of observations $6(188) + (558) = 1686$. Thus the approximation (4.7) which gives $6(187.6) + 634 = 1760$ is less than 5% increase over the more exact result 1686 .

For Model 2, if $\underline{p} = .15$, $\overline{p} = .30$ and $P_1^* = .90$ then $\lambda_1 = 1.6450$, $\lambda_2 = -2.1282$, $\Delta = \overline{p} - \underline{p} = .15$ and the initial guess for k by (4.13) is $k_0 = 3.3$, although we used $k_0 = 3$. By (4.7) with $k_0 = 3$ the initial guess for N is 145 and by (4.10) $c = .106$. Then the value of c hardly changes at all and after some searching, with the help of (4.13) and repeated calculations of (4.4), we obtain $N = 135$, $c = .1061$ and $k = 3.548$. This results in $P_1 = .9000$ and in a total number of observations equal to $6(135) + 479 = 1289$. Thus the approximate solution (4.7) with $k = 3$ which gives $6(145) + 435 = 1305$ is a little more than a 1% increase over the more exact result 1289 for getting (4.4) equal to or greater than $P_1^* = .90$.

For the original condition involving the double integral (3.10), we also want to consider two cases according as we use $k = 3$ or as we find the best possible value of k. For $k = 3$ the best result is obtained at $N = 138$ and $c = .104$ which gives for the total result $6(135) + 415 = 1243$.

The exact solution of the double integral in (3.10) gives the result $N = 130$, $c = .104$ and $k = 3.53$. This yields $(6+k)N = 1239$, which is 4% less than the approximate answer 1289 based on (4.4).

It should be noted that both of the approximations here (i.e., the use of (4.4) instead of (3.10) and the use of $k = 3$ instead of the best value of k) are conservative in the sense that they yield a value of $P_1 > P_1^*$ by asking for a slightly larger value of $(6+k)N$ than is required.

It is easy to show that the condition (1.8) is also satisfied in our illustrative example for any reasonable value of P_0^* and in particular for $P_0^* = .99$. In fact, since $I_{k'}(h, \rho')$ in (1.11) is monotonically increasing in ρ and since ρ' in (3.5) is monotonically decreasing in k, we can set $\rho' = .6$, $k = 5$ in (1.11), $k = 3$ on the right side of (3.5) and obtain as a lower bound

$$(4.15) \qquad P_0 > I_5(c\sqrt{3N/2pq}\ ,\ .6) = I_5(4.145,\ .6) = .99993.$$

Hence P_0 is very close to one and, in particular, is greater than $P_0^* = .99$.

On the basis of the above result it is claimed that certain decisions can be added to CD_1 in (1.8) and (1.9) and to CD_2 in (1.10) as suggested in Remark 2 of Section 1 above with only a negligible change in the joint probability level. These details will not be included in this paper.

The relationship between the results for Model 1 and Model 2 is now fairly clear, at least for the criterion $P_1 \geq P_1^*$. If we replace the numerical value of p_a by the value of $\bar{p}$ and Δ' by Δ in (3.10) then the same expression gives P_1 for Model 2. This gives a new $\bar{p}$ which is larger and makes the problem easier, i.e., the value of N and of $(6+k)N$ required is smaller. This relationship also is the main reason for including both models in the same paper.

5. Approximate Solution for Equality in (1.10).

Using the same methods that were used above to obtain (4.4), we can also approximate the single integral for P_2 in (3.12). We assume that N is large and that c' is such that $0 < c' < \Delta$. Then, using (4.1) and (4.2) with 3 replaced by 5, we first obtain for the integrand in (3.12)

$$(5.1) \qquad [1 - \Phi(A)]\Phi^5(B) = [1 - \Phi(A)][1 - \Phi(-B)]^5 \sim 1 - \Phi(A) - 5\Phi(-B)$$

$$= \Phi(-A) - 5\Phi(-B).$$

Then the integral in (3.12) is approximated by

$$(5.2) \qquad P_2 \sim \int_{-\infty}^{\infty} [\Phi\left(\frac{-y\sqrt{\underline{pq}/k} + (\Delta - c')\sqrt{N}}{\sqrt{\overline{pq}}}\right) - 5\Phi\left(\frac{-y\sqrt{\underline{pq}/k} - c'\sqrt{N}}{\sqrt{\underline{pq}}}\right)]\varphi(y)dy$$

$$= \Phi\left(\frac{(\Delta - c')\sqrt{N}}{\sqrt{\overline{pq} + \underline{pq}/k}}\right) - 5\Phi\left(\frac{-c'\sqrt{N}}{\sqrt{\underline{pq}(1+1/k)}}\right).$$

First we wish to find the value of c' that maximizes this expression for fixed N and k. Then we want to show that with this maximizing value of c' and with the values of k and N that were found from condition (1.9), the condition (1.10) is also satisfied. We will show this only in our example; we do not claim this is always true. If it is not the case, then one option is to use (3.12) or (5.2) to obtain values of N, k and c' and show that (3.10) or (4.4) is automatically satisfied when the optimal value of c is used.

A differentiation of (5.2) with respect to c' gives us a quadratic equation

$$(5.3) \qquad \frac{(c')^2}{\underline{pq}(1+1/k)} - \frac{(c'-\Delta)^2}{\overline{pq} + \underline{pq}/k} - \frac{2}{N}\ln\left(5\sqrt{\frac{\overline{pq}+\underline{pq}/k}{\underline{pq}(1+1/k)}}\right) = 0 .$$

From assumption (1.7) it follows that the coefficient of $(c')^2$ in (5.3) is positive. Since the value of the quadratic is negative at $c' = 0$ it

follows that we want the larger of the two roots (with a plus sign in front of the radical). The second derivative is easily seen to be negative for all c' in the interval $(0, \Delta)$ and hence this largest root of (5.3) must yield a maximum; we do not write out the explicit solution of the quadratic in (5.3).

In our numerical example for Model 1 we take $N = 188$, $\underline{p} = \Delta = .15$ and $k = 2.97$ and obtain results for c' and P_2 in the configuration ω_2. From (5.3) the value of c' that maximizes P_2 is .081 and by (5.2) we obtain the approximation $P_2 \sim .952$. Thus we note that the condition (1.10) for $P_2^* = .75$ is satisfied under Model 1.

In our numerical example for Model 2 we take $N = 130$, $\underline{p} = \Delta = .15$ and $k = 3.53$ and obtain the following results for c' and P_2 in the configuration ω_2. From (5.3) the value of c' that maximizes P_2 is .0854 and by (5.2) this yields the approximation $P_2 \sim .8912$. Thus we note that condition (1.10) for $P_2^* = .75$ is satisfied under both models.

If we had set $P_1^* = P_2^* = .90$ then we note above at $N = 130$, $k = 3.53$, $c' = .0854$ under Model 2 that P_2 is only .8912 and hence we need a larger N-value to reach $P_2^* = .9$ and also a larger value for $(6+k)N$. Using the approximation (5.2) we find that under Model 2 with $c' = .0849$ and $k = 3.53$ we need $N = 136$ to satisfy both (1.9) and (1.10). This yields a total of $(6+k)N = 1299$ observations. This does not mean that the requirement (1.10) is necessarily always harder to satisfy than (1.9). The question is what total is obtained if we make (1.9) the main requirement and start with the k-value that does the best job for (1.10); we now start to do that by first finding the best k-value for (1.10).

Using (5.2) to find $\partial N/\partial k$ (by treating c' as a fixed constant) we substitute this in (4.11) and obtain

$$(5.4)\qquad k = \left\{ \frac{pq(1+\frac{6}{k})}{\theta_1^2} \left[\frac{\frac{(\Delta-c')}{\theta_1}\varphi\left(\frac{(\Delta-c')\sqrt{N}}{\theta_1}\right) + 5\left(\frac{\theta_1}{\theta_2}\right)^2 \frac{c'}{\theta_2}\varphi\left(\frac{-c'\sqrt{N}}{\theta_2}\right)}{\left(\frac{\Delta-c'}{\theta_1}\right)\varphi\left(\frac{(\Delta-c')\sqrt{N}}{\theta_1}\right) + 5\left(\frac{c'}{\theta_2}\right)\varphi\left(\frac{-c'\sqrt{N}}{\theta_2}\right)} \right] \right\}^{\frac{1}{3}}$$

where θ_1 and θ_2 are respectively the two radicals in the final expression for (5.2). A good initial guess for k to solve (5.4) by iteration is obtained by setting the entire square bracket in (5.4) equal to one, since $\varphi\left(\frac{-c'\sqrt{N}}{\theta_2}\right)$ is much smaller than $\varphi((\Delta-c')\sqrt{N}/\theta_1)$. This gives the simpler result

$$(5.5)\qquad k \sim \left[\frac{pq(1+\frac{6}{k})}{\theta_1^2} \right]^{1/3},$$

which also needs to be iterated. For Model 2 we obtain $k = 1.322$. By (5.3) and (5.2) we then find that we need $c' = .086$ and $N = 178$. Remarkably enough, this yields $(6+k)N = 1304$ observations, which is very close to the 1299 obtained above. However by checking the requirement (1.9) at this value of k ($k = 1.322$) we find from (4.9) that the best c is .1087 and by (4.4) this only yields $P_1 = .8488$. Hence even for $P_1^* = P_2^* = .90$ we still need to use the requirement (1.9) to determine k, N and c. As mentioned above the approximate solution in this case is $N = 136$, $k = 3.53$, $c = .1061$, $c' = .0849$ and this yields $(6+k)N = 1299$ observations. Using the more exact integral expression in (3.12) we can expect at most a 5% reduction in this result for $(6+k)N$; exact results based on normal theory have been computed (cf. Table 1).

In summary, we used requirement (1.9) to determine k, c and an initial estimate of N. Then requirement (1.10) determined c' and helped to give a revised estimate of N. In our first example when $P_2^* = .75$ there was no revision needed but when $P_2^* = .90$ under Model 2 we revised N from 130 to 136. Requirement (1.8) was automatically satisfied and was never really used.

6. Combinations Without a Control.

One of the reasons why our sample sizes in Section 5 are large is that we include a control with each set of four chemicals. Clearly we build up information on the control as we go along with different sets of 4 chemicals and after some time it is possible that the control can be dispensed with altogether. It is therefore of interest to get the corresponding sample sizes when the controls are eliminated. The techniques and notation here are quite similar to that in Section 5 and we omit many details.

Under both Models 1 and 2

$$(6.1)\qquad P_0 = P\{X_i < 3Nc_0 \ (i = 1,2,3,4)\,|\,\omega_0\}$$

$$= P\{Y_i < (c_0 - \underline{p})\sqrt{\frac{3N}{\underline{pq}}} \ (i = 1,2,3,4)\}$$

$$= P\{Z_i < Z_0\sqrt{\frac{\rho}{1-\rho}} + (c_0 - \underline{p})\sqrt{\frac{3N}{(1-\rho)\underline{pq}}} \ (i = 1,2,3,4)\}$$

$$= \int_{-\infty}^{\infty} \Phi^4\left(\frac{z\sqrt{\rho} + h}{\sqrt{1-\rho}}\right) d\Phi(z),$$

where $\rho = 1/3$ and $h = (c_0 - \underline{p})\sqrt{3N/\underline{pq}}$.

Under Model 1 (the same result holds for Model 2 if we replace p_a everywhere in (6.2) and (6.3) by $\bar{p}$), letting $p^* = (2\underline{p} + p_a)/3$,

$$(6.2)\qquad P_1 = P\{X_4 > 3Nc_0,\ X_i < 3Nc_0 \ (i = 1,2,3)\,|\,\omega_1\}$$

$$= P\{Y_4 > (c_0 - p_a)\sqrt{\frac{3N}{p_a q_a}} \ , \ Y_i < \frac{3(c_0 - p^*)\sqrt{N}}{\sqrt{2\underline{pq} + p_a q_a}} \ (i = 1,2,3)\}$$

$$= P\{Z_4 > Z_0\sqrt{\frac{\rho^*}{1-\rho^*}} + \frac{(c_0 - p_a)\sqrt{3N}}{\sqrt{(1-\rho^*)p_a q_a}} \ , \ Z_i < Z_0\sqrt{\frac{\rho}{1-\rho}} + \frac{3(c_0 - p^*)\sqrt{\ }}{\sqrt{(1-\rho)(2\underline{pq}+p_a q}}$$

$$(i = 1,2,3)\}$$

$$= \int_{-\infty}^{\infty} [1-\Phi\left(\frac{zp_a q_a - 3(p_a - c_0)\sqrt{N\underline{pq}}}{\sqrt{(3\underline{pq} - p_a q_a)p_a q_a}}\right)]\Phi^3\left(\frac{z\sqrt{\underline{pq}} + 3(c_0 - p^*)\sqrt{N}}{\sqrt{\underline{pq} + p_a q_a}}\right) d\Phi(\ )$$

where $p^* < c_0 < p_a$ and in (6.2) we used the quantities

$$(6.3) \qquad \rho = \rho_{12} = \frac{\underline{pq}}{2\underline{pq} + p_a q_a} \text{ and } \rho^* = \frac{\rho_{14}^2}{\rho} = \frac{p_a q_a}{3\underline{pq}} .$$

To obtain P_1 for Model 2 we replace p_a by $\bar{p}$ and q_a by $\bar{q}$ in (6.2) and (6.3) and in the limitations on c_0.

Under both Model 1 and Model 2 we introduce a new constant c_0' for deciding whether the mixtures are bad and obtain

$$(6.4) \qquad P_2 = P\{X_{3,4} > Nc_0', X_{i,j} < Nc_0' \text{ for } (i,j) \neq (3,4) | \omega_2\}$$

$$= P\{Y_{3,4} > \frac{-(\bar{p} - c_0')\sqrt{N}}{\sqrt{\bar{p}\bar{q}}}, Y_{i,j} < \frac{(c_0' - \underline{p})\sqrt{N}}{\sqrt{\underline{pq}}} \text{ for } (i,j) \neq (3,4)\}$$

$$= \Phi\left(\frac{(\bar{p} - c_0')\sqrt{N}}{\sqrt{\bar{p}\bar{q}}}\right)\Phi^5\left(\frac{(c_0' - \underline{p})\sqrt{N}}{\sqrt{\underline{pq}}}\right),$$

where $\underline{p} < c_0' < \bar{p}$. Note that this result does not depend on p_a.

As in Section 5 our main interest is to determine N and c_0 by setting $P_1 = P_1^*$. We determine c_0' by maximizing P_2 and reporting the resulting values of P_2 and P_0. Such values must exist since for $p^* < c_0 < p_a$ the value of P_1 approaches 1 as $N \to \infty$ and we note that for $\underline{p} < c_0' < \bar{p}$ the same is true for P_2.

Numerical results for these cases are compared with those of Section 5 and the one-at-a-time model in Tables 1 and 2.

The corresponding approximation formulas for (6.2) using the same technique as in Section 4 is

$$(6.5) \qquad P_1 \sim \Phi\left(\frac{(p_a - c_0)\sqrt{3N}}{\sqrt{p_a q_a}}\right) - 3\Phi\left(\frac{-3(c_0 - p^*)\sqrt{N}}{\sqrt{2\underline{pq} + p_a q_a}}\right).$$

By setting the first term equal to $(1+P^*)/2$ and the second equal to $(1-P^*)/2$, we obtain as a first approximation

$$(6.6) \qquad N \sim \left[\frac{\lambda_1\sqrt{3p_aq_a} - \lambda_2\sqrt{2\underline{pq} + p_aq_a}}{3(p_a - p^*)} \right]^2,$$

$$(6.7) \qquad c_0 \sim \frac{p^*\lambda_1\sqrt{3p_aq_a} - p_a\lambda_2\sqrt{2\underline{pq} + p_aq_a}}{\lambda_1\sqrt{3p_aq_a} - \lambda_2\sqrt{2\underline{pq} + p_aq_a}},$$

where λ_1 and λ_2 are the standard normal percentiles corresponding to $(1 + P^*)/2$ and $(1 - P^*)/6$, respectively. A better approximation for c_0, obtained by differentiating (6.5) with respect to c_0 to find a maximum, is the solution in c of the quadratic

$$(6.8) \qquad \frac{3(c - p^*)^2}{2\underline{pq} + p_aq_a} - \frac{(p_a - c)^2}{p_aq_a} - \frac{1}{3N} \ln\left(\frac{27p_aq_a}{2\underline{pq} + p_aq_a} \right) = 0.$$

7. One-at-a-Time Procedures With and Without Controls.

Since we are experimenting with pairs of chemicals there should be advantages over one-at-a-time testing. We do not save on the total number of observations (to attain the same probability of a correct classification equal to or greater than .9) over the one-at-a-time method. In fact, there is approximately a 50% increase since the one-at-a-time method with a control requires $5(183.2) = 916$ if we take $k = N_0/N_1 = 1$ and it requires $4(147.5) + 26 = 854$ if we take $k = 1.785$; as noted above the combination scheme requires a total of 1239 observations. One reason for this is that we have restrained our combination procedure to consider all $\binom{4}{2} = 6$ pairs of treatments simultaneously instead of doing them sequentially. In addition, as already pointed out, the one-at-a-time procedures gives us no information at all about the results of combining chemicals.

We now give corresponding expressions for P_0 and P_1 using the one-at-a-time procedure, without giving all the details. Since the chemicals are not combined, we make no reference to either Model 1 or 2 and we omit P_2, but the normal approximation to the binomial is again used.

For the one-at-a-time procedure with a control

$$(7.1)\qquad P_0 = P\{\frac{X_i}{N} - \frac{X_0}{kN} < c \ (i = 1,2,3,4)|\omega_0\}$$

$$= P\{Y_i < Y_0\sqrt{1/k} + c\sqrt{\frac{N}{\underline{pq}}} \ (i = 1,2,3,4)\} = I_5(h, \rho'),$$

where

$$(7.2)\qquad \rho' = \frac{1}{k+1} \quad \text{and} \quad h = c\sqrt{\frac{Nk}{\underline{pq}(k+1)}} \ .$$

For the one-at-a-time procedure with a control

$$(7.3)\qquad P_1 = P\{\frac{X_4}{N} - \frac{X_0}{kN} > c, \ \frac{X_i}{N} - \frac{X_0}{kN} < c \ (i = 1,2,3)|\omega_1\}$$

$$= P\{Y_4 > Y_0\sqrt{\frac{\underline{pq}}{k\bar{p}\bar{q}}} - \frac{(\Delta-c)\sqrt{N}}{\sqrt{\bar{p}\bar{q}}} \ , \ Y_i < Y_0\sqrt{1/k} + c\sqrt{\frac{N}{\underline{pq}}} \ (i = 1,2,3)\}$$

$$= \int_{-\infty}^{\infty} [1-\bar{\Phi}(\frac{y\sqrt{\underline{pq}} - (\Delta-c)\sqrt{Nk}}{\sqrt{k\bar{p}\bar{q}}})]\bar{\Phi}^3(y\sqrt{1/k} + c\sqrt{\frac{N}{\underline{pq}}})d\bar{\Phi}(y),$$

where $0 < c < \Delta = \bar{p} - \underline{p}$. For c in this range we note that $P_1 \to 1$ as $N \to \infty$.

For the one-at-a-time procedure without a control

$$(7.4)\qquad P_0 = P\{\frac{X_i}{N} < c_0 \ (i = 1,2,3,4)|\omega_0\}$$

$$= P\{Y_i < \frac{(c_0 - \underline{p})\sqrt{N}}{\sqrt{\underline{pq}}} \ (i = 1,2,3,4)\} = \bar{\Phi}^4(\frac{(c_0 - \underline{p})\sqrt{N}}{\sqrt{\underline{pq}}}),$$

where $\underline{p} < c_0 < \bar{p}$.

For the one-at-a-time procedure without a control

$$(7.5)\qquad P_1 = P\{\frac{X_4}{N} > c_0, \ \frac{X_i}{N} < c_0 \ (i = 1,2,3)|\omega_1\}$$

$$= P\{Y_4 > -\frac{(\bar{p} - c_0)\sqrt{N}}{\sqrt{\bar{p}\bar{q}}} \ , \ Y_i < \frac{(c_0 - \underline{p})\sqrt{N}}{\sqrt{\underline{pq}}} \ (i = 1,2,3)\}$$

$$= \bar{\Phi}(\frac{(\bar{p} - c_0)\sqrt{N}}{\sqrt{\bar{p}\bar{q}}})\bar{\Phi}^3(\frac{(c_0 - \underline{p})\sqrt{N}}{\sqrt{\underline{pq}}})$$

for $\underline{p} < c_0 < \bar{p}$. For c in this range we note that $P_1 \to 1$ as $N \to \infty$.

The approximation formula for (7.3) analogous to (6.5) is

$$(7.6) \qquad P_1 \sim \Phi\left(\frac{(\Delta - c)\sqrt{N}}{\sqrt{\bar{p}\bar{q} + \underline{p}\underline{q}/k}} \right) - 3\Phi\left(\frac{-c\sqrt{N}}{\sqrt{(k+1)\underline{p}\underline{q}/k}} \right).$$

By setting the first term equal to $(1+P^*)/2$ and the second equal to $(1-P^*)/2$, we obtain as a first approximation

$$(7.7) \qquad N \sim \left[\frac{\lambda_1\sqrt{\bar{p}\bar{q} + \underline{p}\underline{q}/k} - \lambda_2\sqrt{(k+1)\underline{p}\underline{q}/k}}{\Delta} \right]^2$$

$$(7.8) \qquad c \sim \frac{-\Delta\lambda_2\sqrt{(k+1)\underline{p}\underline{q}/k}}{\lambda_1\sqrt{\bar{p}\bar{q} + \underline{p}\underline{q}/k} - \lambda_2\sqrt{(k+1)\underline{p}\underline{q}/k}}$$

where λ_1 and λ_2 are the same as in (6.7) and in Section 4. A better approximation for c, obtained by differentiating (7.6) with respect to c to find a maximum, is the solution in c of the quadratic

$$(7.9) \qquad \frac{kc^2}{(k+1)\underline{p}\underline{q}} - \frac{(\Delta-c)^2}{\bar{p}\bar{q} + \underline{p}\underline{q}/k} - \frac{1}{N} \ln\left(\frac{9(\bar{p}\bar{q} + \underline{p}\underline{q}/k)}{(k+1)\underline{p}\underline{q}/k} \right) = 0.$$

8. Acknowledgement.

The authors wish to acknowledge with thanks the many encouraging conversations we had with Professor Robert Elashoff of the University of California at San Francisco.

REFERENCES

[1] Bliss, C. I. (1939). Toxicity of poisons applied jointly. Ann. Appl. Biol. 26 585-615.

[2] Finney, D. J. (1971). Probit Analysis. (Chap. XI-3rd ed.) Cambridge University Press.

[3] Plackett, R. L. and Hewlett, P. S. (1967). A comparison of two approaches to the construction of models for quantal responses to mixtures of drugs. Biometrics 23 27-44.

TABLE 1. Comparisons With a Control ($\underline{p} = .15$, $\overline{p} = .30$).

P_1^*	Model 1 $T = (6+k)N$	Model 2 $T = (6+k)N$	1-at-a-time $T = (4+k)N$
.75	$N = 105$; $k = 2.81$ $c = .092$; $T = 925$ $c' = .091$; $R_1 = .572$	$N = 72$; $k = 3.41$ $c = .108$; $T = 678$ $c' = .100$; $R_2 = .780$	$N = 86$; $k = 2.15$ $c = .088$; $T = 529$ ----
.90	$N = 188$; $k = 2.97$ $c = .089$; $T = 1686$ $c' = .081$; $R_1 = .581$	$N = 130$; $k = 3.53$ $c = .104$; $T = 1239$ $c' = .0855$; $R_2 = .791$	$N = 162$; $k = 2.05$ $c = .078$; $T = 980$ ----
.95	$N = 259$; $k = 2.88$ $c = .088$; $T = 2300$ $c' = .077$; $R_1 = .581$	$N = 179$; $k = 3.45$ $c = .103$; $T = 1692$ $c' = .081$; $R_2 = .790$	$N = 219$; $k = 2.10$ $c = .072$; $T = 1336$ ----

TABLE 2. Comparisons Without a Control ($\underline{p} = .15$, $\overline{p} = .30$).

P_1^*	Model 1 $T = (6+k)N$	Model 2 $T = (6+k)N$	1-at-a-time $T = (4+k)N$
.75	$N = 62$; --- $c = .244$; $T = 372$ $c' = .244$; $R_1 = .581$	$N = 46$; --- $c = .2610$; $T = 276$ $c' = .252$; $R_2 = .783$	$N = 54$; --- $c = .239$; $T = 216$ ----
.90	$N = 111$; --- $c = .239$; $T = 666$ $c' = .233$; $R_1 = .595$	$N = 85$; --- $c = .251$; $T = 510$ $c' = .237$; $R_2 = .776$	$N = 99$; --- $c = .2295$; $T = 396$ ----
.95	$N = 150$; --- $c = .2374$; $T = 900$ $c' = .229$; $R_1 = .600$	$N = 112$; --- $c = .2526$; $T = 672$ $c' = .233$; $R_2 = .804$	$N = 135$; --- $c = .226$; $T = 540$ ----

1. The N-value is the number of animals needed for each cell (or combination of chemicals) except for control groups where we use kN. The constant c is the critical value needed to make the procedure explicit; the value of c' that maximizes P_2 is also given in all cases. The value T is the total number of observations including the control. R_i is the ratio of the required one-at-a-time T-value to the corresponding required T-value for Model i ($i = 1, 2$).

2. It is interesting to note that i) the values of N; T, c and c' are all monotonic in P_1^* and ii) the values of k are erratic, due to the number-theoretic aspect of minimizing $(6+k)N$ or $(4+k)N$, but are insensitive to changes in P_1^* and are approximately constant in each column of Table 1.